



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :		A1	(11) International Publication Number: WO 99/28494
C12Q 1/06, C12P 19/34, C07H 21/00			(43) International Publication Date: 10 June 1999 (10.06.99)
(21) International Application Number: PCT/US98/25664		02142 (US). VAN CAUTER, Staf [BE/US]; 145 Tuttles Point Road, Guilford, CT 06437 (US).	
(22) International Filing Date: 3 December 1998 (03.12.98)		(74) Agent: MYERS, P., Louis; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).	
(30) Priority Data:		60/067,517 4 December 1997 (04.12.97) US 60/069,642 15 December 1997 (15.12.97) US 60/071,443 14 January 1998 (14.01.98) US	
(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Applications		US 60/067,517 (CON) Filed on 4 December 1997 (04.12.97) US 60/069,642 (CON) Filed on 15 December 1997 (15.12.97) US 60/071,443 (CON) Filed on 14 January 1998 (14.01.98)	
(71) Applicant (for all designated States except US): PACKARD BIOSCIENCE COMPANY [US/US]; 800 Research Parkway, Meriden, CT 06450 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(72) Inventors; and		Published	
(75) Inventors/Applicants (for US only): TAYLOR, Seth [US/US]; Apartment 2-9B, 100 Memorial Drive, Cambridge, MA		With international search report.	

(54) Title: METHODS OF USING PROBES FOR ANALYZING POLYNUCLEOTIDE SEQUENCE

## (57) Abstract

A method of detecting the presence of, or identifying, a genetic event comprising: providing a target nucleic acid having a genetic event; providing a plurality of nucleotide probes wherein each probe: (a) is complementary to a region immediately 5' to the site of the genetic event and complementary to a region immediately 3' to the site of the genetic event, (b) has an identification region, e.g., a 5' terminal region or a 3' terminal region, which identification region is not repeated in another probe in the plurality, i.e., it is unique in the plurality, and is complementary to the target sequence, and (c) wherein the nucleotide at the position corresponding to the site of the first genetic event can be chosen from A, T, G or C, wherein the plurality of nucleotide probes includes a first probe having a first base at the position corresponding to the site of the genetic event and a second probe having a second (i.e., different from the first base) base at the position corresponding to the site of the genetic event; contacting the target with the plurality of probes under conditions wherein only probes having complementarity with the target nucleic acid will hybridize; thereby analysing the target nucleic acid.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Larvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## METHODS OF USING PROBES FOR ANALYZING POLYNUCLEOTIDE SEQUENCE

5

## BACKGROUND OF THE INVENTION

The invention relates to a method of analyzing a polynucleotide sequence.

## SUMMARY OF THE INVENTION

10

In general, the invention features, a method of analyzing a polynucleotide sequence in a sample. The method includes:

providing a sample which includes a polynucleotide sequence to be analyzed and at least one extraneous nucleic acid molecule;

15

providing a plurality of selector probes wherein each selector probe is complementary to a region of the polynucleotide sequence to be analyzed and is not complementary to the extraneous nucleic acid (selector probes can be complementary to different regions of the poly-nucleotide sequence),

20

contacting the sample with the plurality of selector probes under conditions which allow hybridization of the polynucleotide sequence to be analyzed to a selector probe but which do not allow hybridization of the extraneous nucleic acid to hybridize to a selector probe;

25

allowing a nucleic acid molecule to be analyzed to hybridize with a selector probe;

selecting a product of that hybridization, e.g., the selector probe, the nucleic acid to be analyzed, or both;

30

providing an array of a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region complementary to the plurality of selector probes;

hybridizing the selected nucleic acid molecule (which does not include the extraneous nucleic acid molecule) with the array of capture probes, thereby detecting or identifying a selected nucleic acid molecule which bound to the polynucleotide sequence and thereby analyzing the polynucleotide sequence.

## 2.

In preferred embodiments the polynucleotide sequence is: a DNA molecule: all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly a human genomic fragment; a cDNA, particularly a human cDNA.

5 In preferred embodiments the polynucleotide sequence is: an RNA molecule: nucleic acids derived from RNA transcripts; wild type RNA; mutant RNA, particularly a human RNA.

10 In preferred embodiments the polynucleotide sequence is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

In preferred embodiments, the target is anchored to a substrate and the addition of selector probes to the anchored target is repeated at least one time to increase the number of selector probes.

15

In preferred embodiments, the selector probes are anchored to a substrate and the addition of target the anchored probes is repeated at least one time to increase the number of target molecules.

20

In preferred embodiments the polynucleotide sequence is coupled to a support prior to hybridizing with the selector probes.

In preferred embodiments the selector probes are coupled to a support prior to hybridizing with the target.

25

In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

In preferred embodiments the selector probes are selected to be to maximize uniqueness, i.e., to minimize the number of non-unique selector probes. This can be done with computer-aided methods. This can be done prior to dividing the probes into pools.

## 3.

- e.g., a computer-aided method can be used to select or locate a group of selector probes in which the frequency of repeated complementarity regions is minimized or reduced to zero. If there are no repeats there need be only one pool. If there is non-uniqueness, i.e., if a complementarity region is repeated, uniqueness can be obtained by placing the selector probes having the same complementarity region in separate pools.

In preferred embodiments the selector probes are labeled so that hybridization to the array of capture probes can be followed. The label can be a fluorescent label. The selector probes of a first pool can be labeled with a first label and the selector probes of a second pool can be labeled with a second label. If only one label is used then pools must be hybridized separately to the array of capture probes.

In preferred embodiments the capture probes are single stranded probes in an array.

15 In preferred embodiments the capture probes have a structure comprising a double stranded portion and a single stranded portion in an array.

In preferred embodiments hybridization is detected by mass spectrophotometry, e.g., by MALDI-TOF mass spectrophotometry.

20 In preferred embodiments probes are selected for minimal cross-hybridization with other probes.

In preferred embodiments the target or the selector probe has attached thereto a first member of a proximity detector pair and hybridization to the array allows the first member to be brought into proximity with a second member to provide a signal.

In another aspect, the invention features, a method of detecting a target sequence in a sample. The method includes:

30 providing a sample containing a target sequence;

providing a plurality of selector probes wherein each selector probe is complementary to the target sequence,

## 4.

contacting the target sequence with the plurality of selector probes under conditions wherein only selector probes having complete complementarity with the target sequence will hybridize;

selecting selector probes that hybridize to the target sequence;

5 providing an array of a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region complementary to the plurality of selector probes;

10 hybridizing the selector probes which hybridize to the target sequence to the array of capture probes, thereby detecting or identifying a selector probe which bound to the target sequence and thereby detecting said target sequence in a sample.

In preferred embodiments the target sequence is: a DNA molecule: all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly a human 15 genomic fragment; a cDNA, particularly a human cDNA.

In preferred embodiments the target sequence is: an RNA molecule: nucleic acids derived from RNA transcripts; wild type RNA; mutant RNA, particularly a human RNA.

20 In preferred embodiments, the target is anchored to a substrate and the addition of selector probes to the anchored target is repeated at least one time to increase the number of selector probes.

25 In preferred embodiments, the selector probes are anchored to a substrate and the addition of target the anchored probes is repeated at least one time to increase the number of target molecules.

In preferred embodiments the selector probes are coupled to a support prior to 30 hybridizing with the target.

In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other

## 5.

treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

In preferred embodiments the target sequence is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

In preferred embodiments the selector probes are selected to be to maximize uniqueness, i.e., to minimize the number of non-unique selector probes. This can be done with computer-aided methods. This can be done prior to dividing the probes into pools.

10 e.g., a computer-aided method can be used to select or locate a group of selector probes in which the frequency of repeated complementarity regions is minimized or reduced to zero. If there are no repeats there need be only one pool. If there is non-uniqueness, i.e., if a complementarity region is repeated, uniqueness can be obtained by placing the selector probes having the same complementarity region in separate pools.

15

In preferred embodiments the selector probes are labeled, e.g., so that hybridization to the array of capture probes can be followed. The label can be a fluorescent label. The selector probes of a first pool can be labeled with a first label and the selector probes of a second pool can be labeled with a second label. If only one label is used then pools must be 20 hybridized separately to the array of capture probes.

In preferred embodiments the capture probes are single stranded probes in an array.

25 In preferred embodiments the capture probes have a structure comprising a double stranded portion and a single stranded portion in an array.

In preferred embodiments hybridization is detected by mass spectrophotometry, e.g., by MALDI-TOF mass spectrophotometry.

30 In preferred embodiments probes are selected for minimal crosshybridization with other probes.

## 6.

In preferred embodiments the target or the selector probe has attached thereto a first member of a proximity detector pair and hybridization to the array allows the first member to be brought into proximity with a second member to provide a signal.

5 In another aspect, the invention features, a method of analysing a target nucleic acid, e.g., of detecting the presence of, or identifying, a genetic event, e.g., a polymorphism, in a target nucleic acid, e.g., a DNA. The method includes:

providing a target nucleic acid having a genetic event;

providing a plurality of nucleotide probes wherein each probe:

10 (a) is complementary to a region immediately 5' to the site of the genetic event and complementary to a region immediately 3' to the site of the genetic event,

(b) has an identification region, e.g., a 5' terminal region or a 3' terminal region, which identification region is not repeated in another probe in the

15 plurality, i.e., it is unique in the plurality, and is complementary to the target sequence, and

(c) wherein the nucleotide at the position corresponding to the site of the genetic event can be chosen from A, T, G, or C,

(optionally) wherein the plurality of nucleotide probes includes a first probe having a

20 first base at the position corresponding to the site of the genetic event and a second probe having a second (i.e., different from the first base) base at the position corresponding to the site of the genetic event;

contacting the target with the plurality of probes under conditions wherein only probes having complementarity with the target nucleic acid will hybridize;

25 thereby analysing the target nucleic acid. The identification region can allow identification of the identity of a base at the genetic event.

In a particularly preferred embodiment the method includes:

30 providing an array of a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region complementary to the identification region of a probe of the plurality of nucleotide probes;

hybridizing the nucleotide probes which hybridize to the target to the array of capture probes, thereby detecting or identifying a nucleotide probe which bound to the target and thereby identifying the genetic event.

The identification region is a naturally occurring sequence in the target molecule. A probe sequence is chosen such that it corresponds to a part of the target sequence which includes an identification region. The identifier region can be at least 2, 4, 6, 8, 10, or 12 or more base pairs in length. In preferred embodiments it is less than 8, 10, 12, 14, 20, or 30 base pairs in length. In preferred embodiments it is between 2 and 40, 4 and 40, 6 and 30, 6 and 20 or 6 and 10 base pairs long. It can overlap with or be independent of other regions of the probe discussed herein, e.g., the target position, or positions on the 3' or 5' thereof. The length will, depend, in part, on the sequence of the region and on how many distinct identifier regions are required.

In preferred embodiments the target nucleic acid is: a DNA molecule: all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly a human genomic fragment; a cDNA, particularly a human cDNA.

In preferred embodiments the genetic event is: a polymorphism, a single base pair change; a naturally occurring variant; an allele associated with a disease state; a mutation; an environmentally induced lesion.

In preferred embodiments the target nucleic acid is coupled to a support prior to hybridizing with the nucleotide probes.

In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

30

In preferred embodiments the target nucleic acid is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

## 8.

- In preferred embodiments the nucleotide probes are selected to be to maximize uniqueness of the identification regions, i.e., to minimize the number of non-unique nucleotide probes. This can be done with computer-aided methods. This can be done prior to dividing the probes into pools, e.g., a computer-aided method can be used to select or 5 locate a group of nucleotide probes in which the frequency of repeated identification regions is minimized or reduced to zero. If there are no repeats there need be only one pool. If there is non-uniqueness, i.e., if an identification region is repeated, uniqueness can be obtained by placing the nucleotide probes having the same identification region in separate pools.
- 10 In preferred embodiments the nucleotide probes are labeled, e.g., so that hybridization to the array of capture probes can be followed. The label can be a fluorescent label. The nucleotide probes of a first pool can be labeled with a first label and the nucleotide probes of a second pool can be labeled with a second label. If only one label is 15 used then pools must be hybridized separately to the array of capture probes.
- 20 In preferred embodiments the capture probes are single stranded probes in an ordered array.
- 25 In preferred embodiments the capture probes have a structure comprising a double stranded region and a single stranded region in an ordered array.
- 30 In preferred embodiments hybridization is detected by mass spectrophotometry, e.g., by MALDI-TOF mass spectrophotometry.
- In preferred embodiments probes are selected for minimal cross-hybridization with other probes.
- 35 In preferred embodiments the target or the selector probe has attached thereto a first member of a proximity detector pair and hybridization to the array allows the first member to be brought into proximity with a second member to provide a signal.

9.

In preferred embodiments the target or a nucleotide probe has attached thereto a first member of a proximity detector pair and hybridization to the array allows the first member to be brought into proximity with a second member to provide a signal.

5 In preferred embodiments, the target is anchored to a substrate.

In another aspect, the invention features, a method of detecting the presence of, or identifying, one or more genetic events, e.g., one or more polymorphism, in a target nucleic acid, e.g., a DNA. The method includes:

10

(1) providing a target nucleic acid having a first genetic event, e.g., a polymorphism, and a second genetic event, e.g., a polymorphism;

15

(2) providing a first plurality of nucleotide probes wherein each probe:

(a) is complementary to a region immediately 5' to the site of the first genetic event, e.g., a polymorphism, and complementary to a region immediately 3' to the site of the genetic event, e.g., a polymorphism,

(b) has an identification region, e.g., a 5' terminal region or a 3' terminal region, which is complementary to the target sequence, and

20

(c) wherein the nucleotide at the position corresponding to the site of the first genetic event, e.g., polymorphism, can be chosen from A, T, G, or C,

(3) providing a second plurality of nucleotide probes wherein each probe:

25

(a) is complementary to a region immediately 5' to the site of the second genetic event, e.g., a polymorphism, and complementary to a region immediately 3' to the site of the second genetic event, e.g., a polymorphism,

(b) has an identification region, e.g., a 5' terminal region or a 3' terminal region, which is complementary to the target sequence, and

30

(c) wherein the nucleotide at the position corresponding to the site of the second genetic event, e.g., a polymorphism, can be chosen from A, T, G, or C,

10.

(4) optionally, organizing the first and second plurality of probes into one or more pools, such that within a pool, an identification region is not repeated in the pool, i.e., it is unique in the pool, and

5 (5) contacting the target with the first and second plurality of nucleotide probes under conditions wherein only probes having complementarity with the target will hybridize;

thereby analysing the target nucleic acid. The identification region can allow identification of the identity of a base at the genetic event.

10

In a particularly preferred embodiment the method further includes:

(6) providing an array of a plurality of capture probes, wherein each of the plurality of capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique  
15 (i.e., not repeated in another capture probe) region complementary to the identification region of a probe of the first and second plurality of nucleotide probes;

(7) hybridizing nucleotide probes which hybridize to the target to the array, thereby detecting or identifying a nucleotide probe which bound to the target and thereby  
20 identifying the polymorphism, provided that in a hybridization all of the nucleotide probes have a unique identification region, e.g., they are all from a single pool.

The identification region is a naturally occurring sequence in the target molecule. A probe sequence is chosen such that it corresponds to a part of the target sequence which  
25 includes an identification region. The identifier region can be at least 2, 4, 6, 8, 10, or 12 or more base pairs in length. In preferred embodiments it is less than 8, 10, 12, 14, 20, or 30 base pairs in length. In preferred embodiments it is between 2 and 40, 4 and 40, 6 and 30, 6 and 20 or 6 and 10 base pairs long. It can overlap with or be independent of other regions of the probe discussed herein, e.g., the target position, or positions on the 3' or 5' thereof. The  
30 length will, depend, in part, on the sequence of the region and on how many distinct identifier regions are required.

In preferred embodiments there is more than one pool and step 5, step 7, or both, are performed separately for each of a plurality of pools.

11.

The numbering in the method is provided for ease of understanding. As will be appreciated by one skilled in the art any order which will result in the objective of the method can be used, e.g., the order of steps 4 and 5 can be reversed. Similarly, the 5 numbering does not mean that steps must be performed discreetly, e.g., steps 2 and 3 can be combined.

In preferred embodiments the first plurality of nucleotide probes includes a first probe having a first base at the position corresponding to the site of the first polymorphism 10 and a second probe having a second (i.e., different from the first base) base at the position corresponding to the site of the first polymorphism. In preferred embodiments all of A, T, G, and C are represented.

In preferred embodiments the second plurality of nucleotide probes includes a first 15 probe having a first base at the position corresponding to the site of the second polymorphism and a second probe having a second (i.e., different from the first base) base at the position corresponding to the site of the second polymorphism. In preferred embodiments all of A, T, G, and C are represented.

20 In preferred embodiments the probes are selected to be optimized for hybridization to the target, e.g., to the region having a polymorphism.

In preferred embodiments the nucleotide probes are selected to be to maximize uniqueness of the identification regions, i.e., to minimize the number of non-unique 25 nucleotide probes. This can be done with computer-aided methods. This can be done prior to dividing the probes into pools. E.g., a computer-aided method can be used to select or locate a group of nucleotide probes in which the frequency of repeated identification regions is minimized or reduced to zero. If there are no repeats there need be only one pool. If there is non-uniqueness, i.e., if an identification region is repeated, uniqueness can be 30 obtained by placing the nucleotide probes having the same identification region in separate pools.

In preferred embodiments the target nucleic acid is: a DNA molecule: all or part of a known gene; all or part of a known polymorphic gene; wild type DNA; mutant DNA; a

12.

genomic fragment, particularly a human genomic fragment; a cDNA, particularly a human cDNA.

5 In preferred embodiments the polymorphism is: a single base pair change; a naturally occurring variant; an allele associated with a disease state; a mutation; an environmentally induced lesion.

10 In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

15 In preferred embodiments the target nucleic acid is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

20 In preferred embodiments the genetic event is: a polymorphism, a single base pair change; a naturally occurring variant; an allele associated with a disease state; a mutation; an environmentally induced lesion.

In preferred embodiments the target nucleic acid is coupled to a support prior to hybridizing with the nucleotide probes.

25 In preferred embodiments the nucleotide probes are labeled, e.g., so that hybridization to the array of capture probes can be followed. The label can be a fluorescent label. The nucleotide probes of a first pool can be labeled with a first label and the nucleotide probes of a second pool can be labeled with a second label. If only one label is used then pools must be hybridized separately to the array of capture probes.

30 The method is described for use with a first and a second polymorphism. It can also be modified for use in detecting a single polymorphism by deleting the second plurality of probes. Likewise, it can be modified to detect more than two polymorphisms.

## 13.

In preferred embodiments hybridization is detected by mass spectrophotometry, e.g., by MALDI-TOF mass spectrophotometry.

In preferred embodiments the target or a nucleotide probe has attached thereto a first member of a proximity detector pair and hybridization to the array allows the first member to be brought into proximity with a second member to provide a signal.

In another aspect, the invention features, a method of analyzing a target nucleotide sequence in a sample. The method includes:

10. providing a target nucleic acid;
- hybridizing to the target nucleic acid a first probe having attached thereto a first member of a proximity detection pair;
- hybridizing to the target nucleic acid a second probe having attached thereto a second member of a proximity detection pair; and
- 15 determining if a signal is generated by the proximity detection pair.

The signal from the probes hybridized to the target nucleic acid (e.g., a polymorphic variant) can be compared to the signal generated by such probes when hybridized to a reference molecule (e.g., a wild type or non disease state associated sequence). Genetic events, e.g., rearrangements, e.g., inversions, deletions, insertions, or translocations, can alter the distance between the bound probes (or inhibit binding of a probe) and thereby generate a characteristic signal. E.g., a deletion could bring two probes into proximity and thereby cause generation of a signal. An insertion could increase the distance and inhibit production of a signal.

- 25 In preferred embodiments the method further includes providing an array of a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region and hybridizing the target to the array. Preferably, the region of the target which hybridizes to the array does not overlap with the region which binds the first probe or the regions which binds the second probe. Position on the array can provide sequence information.
- 30

## 14.

The target nucleic acid can be DNA, e.g., genomic DNA or cDNA. Target nucleic acid can be fragmented, e.g., by shearing or by enzyme digestion.

In preferred embodiments the target nucleic acid is: a DNA molecule: all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly a human genomic fragment; a cDNA, particularly a human cDNA.

In preferred embodiments the target nucleic acid is: an RNA molecule: nucleic acids derived from RNA transcripts; wild type RNA; mutant RNA, particularly a human RNA.

10

In preferred embodiments the target nucleic acid is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

20

In preferred embodiments the capture probes are single stranded probes in an array.

In preferred embodiments the capture probes have a structure comprising a double stranded portion and a single stranded portion in an array.

25

In preferred embodiments hybridization is detected by mass spectrophotometry, e.g., by MALDI-TOF mass spectrophotometry.

In preferred embodiments probes are selected for minimal crosshybridization with other probes.

30

In one embodiment, the first member of the proximity detection pair absorbs energy at a first wavelength and transmits the absorbed energy to the second member of the proximity detection pair. The second member of the proximity detection pair absorbs the

15.

transmitted energy from the first member of the proximity detection pair and emits the absorbed energy at a second wavelength.

In another embodiment, the first member of the proximity detection pair absorbs  
5 energy at a first wavelength and transmits the absorbed energy to the second member of the proximity detection pair in a homogenous sample. The second member of the proximity detection pair absorbs the transmitted energy from the first member of the proximity detection pair and emits the absorbed energy at a second wavelength in a homogenous sample.

10

In another embodiment, the first member of the proximity detection pair is sensitizer particle that produces singlet oxygen, e.g., phthalocyanine, and the second member of the proximity detection pair is a chemiluminescer particle which reacts with the singlet oxygen to produce chemiluminescence, e.g., an olefin.

15

In one embodiment, the second member of the proximity detection pair is a component of a single stranded capture probes in an array.

In one embodiment, the second member of the proximity detection pair is a  
20 component of the capture probes comprising a double stranded portion and a single stranded portion in a capture array.

In one embodiment, the second member of the proximity detection pair is on the surface of each of a plurality of locations of the capture array.

25

In one embodiment, the second member of the proximity detection pair is in an aqueous solution which is applied to the capture array.

30

In another aspect, the invention features, a method of analyzing a polynucleotide sequence in a sample. The method includes:  
providing a target nucleic acid;  
hybridizing to the target nucleic acid a first probe having attached thereto a first member of a proximity detection pair;

## 16.

- hybridizing to the target nucleic acid a second probe having attached thereto a second member of a proximity detection pair, wherein said second probe hybridizes to a region which contains a genetic event, e.g., a polymorphism, e.g., a single nucleotide polymorphism (an SNP), and wherein hybridization of the second probe is determined by 5 the presence or absence of the genetic event;

determining if a signal is generated by the proximity detection pair.

- In preferred embodiments the method further includes providing an array of a plurality of capture probes, wherein each of the capture probes is positionally 10 distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region and hybridizing the target to the array. Preferably, the region of the target which hybridizes to the array does not overlap with the region which binds the first probe or the regions which binds the second probe. Position on the array can provide 15 sequence information.

The target nucleic acid can be DNA, e.g., genomic DNA or cDNA. Target nucleic acid can be fragmented, e.g., by shearing or by enzyme digestion.

- 20 In preferred embodiments the target nucleic acid is: a DNA molecule: all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly a human genomic fragment; a cDNA, particularly a human cDNA.

- 25 In preferred embodiments the target nucleic acid is: an RNA molecule: nucleic acids derived from RNA transcripts; wild type RNA; mutant RNA, particularly a human RNA.

In preferred embodiments the target nucleic acid is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

- 30 In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

17.

In preferred embodiments the capture probes are single stranded probes in an array.

5 In preferred embodiments the capture probes have a structure comprising a double stranded portion and a single stranded portion in an array.

In preferred embodiments hybridization is detected by mass spectrophotometry, e.g., by MALDI-TOF mass spectrophotometry.

10 In preferred embodiments probes are selected for minimal crosshybridization with other probes.

15 In one embodiment, the first member of the proximity detection pair absorbs energy at a first wavelength and transmits the absorbed energy to the second member of the proximity detection pair. The second member of the proximity detection pair absorbs the transmitted energy from the first member of the proximity detection pair and emits the absorbed energy at a second wavelength.

20 In another embodiment, the first member of the proximity detection pair absorbs energy at a first wavelength and transmits the absorbed energy to the second member of the proximity detection pair in a homogenous sample. The second member of the proximity detection pair absorbs the transmitted energy from the first member of the proximity detection pair and emits the absorbed energy at a second wavelength in a homogenous sample.

25

In another embodiment, the first member of the proximity detection pair is sensitizer particle that produces singlet oxygen, e.g., phthalocyanine, and the second member of the proximity detection pair is a chemiluminescer particle which reacts with the singlet oxygen to produce chemiluminescence, e.g., an olefin.

30

In one embodiment, the second member of the proximity detection pair is a component of a single stranded capture probes in an array.

## 18.

In one embodiment, the second member of the proximity detection pair is a component of the capture probes comprising a double stranded protein and a single stranded portion in a capture array.

- 5 In one embodiment, the second member of the proximity detection pair is on the surface of each of a plurality of locations of the capture array.

In one embodiment, the second member of the proximity detection pair is in an aqueous solution which is applied to the capture array.

10

In another aspect the invention features, a method of analyzing a polynucleotide sequence in a sample. The method includes:

- 15 providing a sample which includes a polynucleotide sequence to be analyzed having a test position or nucleotide;

20 providing a plurality of selector probes which are complementary to a region which contains the test nucleotide, including a first probe which binds if the test nucleotide is a first nucleotide, e.g., an A, and a second probe which binds the region if the test nucleotide is a second nucleotide, e.g., a C (optionally the plurality includes a first probe which binds if the test nucleotide is a first nucleotide, a second probe which binds if the test nucleotide is a second nucleotide, a third probe which binds if the test nucleotide is a third nucleotide, and a fourth probe which binds if the test nucleotide is a fourth nucleotide),

25 contacting the sample with the plurality of selector probes under conditions which allow hybridization of the polynucleotide sequence to be analyzed to a selector probe;

- 30 allowing a nucleic acid molecule to be analyzed to hybridize with a selector probes and selecting probes which hybridize;

providing an array of a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region complementary to the plurality of selector probes;

hybridizing the selected selector probes with the array of capture probes, thereby detecting or identifying a selected nucleic acid molecule which bound to the polynucleotide sequence and thereby analyzing the polynucleotide sequence.

19.

In another aspect, the invention features, a method of analyzing a polynucleotide sequence in a sample. The method includes:

providing a sample which includes a polynucleotide sequence to be analyzed and at least one extraneous nucleic acid molecule (e.g., a cDNA other than the cDNA of interest, in a cDNA library);

providing a plurality of selector probes wherein each selector probe is complementary to a region of the polynucleotide sequence to be analyzed and is not complementary to the extraneous nucleic acid,

10 contacting the sample with the plurality of selector probes under conditions which allow hybridization of the polynucleotide sequence to be analyzed to a selector probe but which do not allow hybridization of the extraneous nucleic acid to a selector probe;

allowing a nucleic acid molecule to be analyzed to hybridize with a selector probe;

selecting the selector probes that hybridise to the polynucleotide sequence;

detecting or identifying the selector probes which bound to the polynucleotide 15 sequence, e.g., by mass spectrometry, and thereby analyzing the polynucleotide sequence.

In preferred embodiments the polynucleotide sequence is: a DNA molecule; all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly a human genomic fragment; a cDNA, particularly a human cDNA.

20

In preferred embodiments the polynucleotide sequence is: an RNA molecule; nucleic acids derived from RNA transcripts; wild type RNA; mutant RNA, particularly a human RNA.

25

In preferred embodiments the polynucleotide sequence is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

30 In preferred embodiments the polynucleotide sequence is coupled to a support prior to hybridizing with the selector probes.

In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other

20.

treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

In preferred embodiments the selector probes are selected to be to maximize uniqueness, i.e., to minimize the number of non-unique selector probes. This can be done with computer-aided methods. This can be done prior to dividing the probes into pools. E.g., a computer-aided method can be used to select or locate a group of selector probes in which the frequency of repeated complementarity regions is minimized or reduced to zero. If there are no repeats there need be only one pool. If there is non-uniqueness, i.e., if a complementarity region is repeated, uniqueness can be obtained by placing the selector probes having the same complementarity region in separate pools.

In preferred embodiments the selector probes are labeled so that hybridization to the array of capture probes can be followed. The selector probes of a first pool can be labeled with a first label and the selector probes of a second pool can be labeled with a second label. If only one label is used then pools must be hybridized separately to the array of capture probes.

In preferred embodiments the capture probes are single stranded probes in an array.

In preferred embodiments the capture probes have a structure comprising a double stranded portion and a single stranded portion in an array.

In preferred embodiments, each capture probe is in a discrete location on a support.

In preferred embodiments, each capture probe is in a discrete location in a microtiter plate, e.g., in the wells of a microtitre plate.

In preferred embodiments, each capture probe is in a gel pad on a support comprising more than one gel pads.

In another aspect, the invention features, a method of analyzing a polynucleotide sequence in a sample using a proximity detection pair. The method includes:

21.

- providing a sample which includes a polynucleotide sequence to be analyzed and at least one extraneous nucleic acid molecule;
- providing a plurality of selector probes wherein each selector probe is complementary to a region of the polynucleotide sequence to be analyzed and is not complementary to the extraneous nucleic acid and has first member of the proximity detection pair,
- contacting the sample with the plurality of selector probes under conditions which allow hybridization of the polynucleotide sequence to be analyzed to a selector probe but which do not allow hybridization of the extraneous nucleic acid to hybridize to a selector probe;
- 10 allowing a nucleic acid molecule to be analyzed to hybridize with a selector probe having the first member of the proximity detection pair;
- selecting a product of that hybridization, e.g., the selector probe, the nucleic acid to be analyzed, or both;
- 15 providing an array of a plurality of capture probes having the second member of the proximity detection pair, wherein each of the capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region complementary to the plurality of selector probes;
- 20 hybridization of the selected nucleic acid molecule having the first member of the proximity detection pair with the array of capture probes having the second member of the proximity detection pair, thereby allowing the first and second binding members of the proximity detection pair to come into close proximity;
- illuminating the array of a plurality of capture probes, wherein the illumination 25 produces a reaction in the first member of the proximity detection pair, the product of the reaction from the first member generates a reaction in the second member of the proximity detection pair;
- detecting the reaction produced from the proximity detection pair, thereby analyzing the polynucleotide sequence.
- 30 In one embodiment, the first member of the proximity detection pair absorbs energy at a first wavelength and transmits the absorbed energy to the second member of the proximity detection pair. The second member of the proximity detection pair absorbs the

22.

transmitted energy from the first member of the proximity detection pair and emits the absorbed energy at a second wavelength.

In another embodiment, the first member of the proximity detection pair absorbs  
5 energy at a first wavelength and transmits the absorbed energy to the second member of the proximity detection pair in a homogenous sample. The second member of the proximity detection pair absorbs the transmitted energy from the first member of the proximity detection pair and emits the absorbed energy at a second wavelength in a homogenous sample.

10

In another embodiment, the first member of the proximity detection pair is sensitizer particle that produces singlet oxygen, e.g., phthalocyanine, and the second member of the proximity detection pair is a chemiluminescer particle which reacts with the singlet oxygen to produce chemiluminescence, e.g., an olefin.

15

In one embodiment, the second member of the proximity detection pair is a component of a single stranded capture probes in an array.

20 In one embodiment, the second member of the proximity detection pair is a component of the capture probes comprising a double stranded protein and a single stranded portion in a capture array.

In one embodiment, the second member of the proximity detection pair is on the surface of each of a plurality of locations of the capture array.

25

In one embodiment, the second member of the proximity detection pair is in an aqueous solution which is applied to the capture array.

30 In another aspect, the invention features a composition of matter which includes one or more collections of the probes described herein.

In another aspect, the invention features a method of analysing a nucleic acid, e.g., detecting the presence of, or identifying, a genetic event, e.g., a SNP, in a target nucleic acid, e.g., a DNA. The method includes:

23.

providing a target nucleic acid having a genetic event;

providing a probe or primer, wherein the primer:

- 5        a) is complementary to a region adjacent to the site of the genetic event, e.g., its 3' end is within 1, 2, 5, 10, 20, 30, or 50 nucleotides of the genetic event, and
- b) has a 3' end capable of serving as a priming site for extension;
- c) is, optionally, labeled;
- 10       d) optionally has an identification region, e.g., a 5' terminal region or a 3' terminal region, e.g., an identification region not repeated in another probe in a plurality of probes contacted with the target, i.e., it is unique in the plurality, and is complementary to the target sequence, and

15       contacting the target with primer under conditions wherein only a primer having complementarity with the target nucleic acid will hybridize;

extending the primer across the genetic event using a primer terminating unit to generate a primer-extension product which terminates with a base which is complementary to a base in the genetic event, or generates a primer-extension product which extends beyond the base in the genetic event if the base in the genetic event is not complementary with the primer terminating unit;

thereby analysing the target.

25       In a preferred embodiment the method includes: extending along the single strand which contains the genetic event with one and preferably with 2, 3, or all 4 labeled chain terminating nucleotides, wherein if more than one labeled chain terminating nucleotide is used each of the chain terminators, e.g., A or C, are distinguishable, such that a chain terminator is incorporated at the position corresponding to the genetic event and can indicate  
30       the presence of the genetic event.

In a preferred embodiment identification region allows identification of the identity of the base at the genetic event.

24.

In a preferred embodiment the method further includes:

providing an array having a plurality of capture probes, wherein each of the capture probes is positionally distinguishable from the other capture probes of the plurality and has a unique variable region (not repeated in another capture probe of the plurality); and

5

hybridizing the base terminated primer-extension product to a capture probe of the array, (preferably the region of the base terminated primer extension product which corresponds to the genetic event hybridizes with the variable region of a capture probe)

10

thereby detecting or identifying a genetic event in a target nucleic acid.

The identification region can allow identification of the identity of a base at the genetic event.

15

The identification region is a naturally occurring sequence in the target molecule. A probe sequence is chosen such that it corresponds to a part of the target sequence which includes an identification region. The identifier region can be at least 2, 4, 6, 8, 10, or 12 or more base pairs in length. In preferred embodiments it is less than 8, 10, 12, 14, 20, or 30 base pairs in length. In preferred embodiments it is between 2 and 40, 4 and 40, 6 and 30, 6 and 20 or 6 and 10 base pairs long. It can overlap with or be independent of other regions of the probe discussed herein, e.g., the target position, or positions on the 3' or 5' thereof. The length will, depend, in part, on the sequence of the region and on how many distinct identifier regions are required.

25

In preferred embodiments the target nucleic acid is: a DNA molecule; all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly a human genomic fragment; a cDNA, particularly a human cDNA.

30

In preferred embodiments the genetic event is: a polymorphism; a single base pair change; a naturally occurring variant; an allele associated with a disease state; a mutation; an environmentally induced lesion.

25.

In preferred embodiments the target nucleic acid is coupled to a support prior to hybridizing with the nucleotide probes.

5 In preferred embodiments the method is performed: on a sample from a human subject; and a sample from a prenatal subject; as part of genetic counseling; to determine if the individual from which the target nucleic acid is taken should receive a drug or other treatment; to diagnose an individual for a disorder or for predisposition to a disorder; to stage a disease or disorder.

10 In preferred embodiments the target nucleic acid is: a human sequence; a non-human sequence, e.g., a mouse, rat, pig, primate.

15 In preferred embodiments the nucleotide probes are selected to be to maximize uniqueness of the identification regions, i.e., to minimize the number of non-unique nucleotide probes. This can be done with computer-aided methods. This can be done prior to dividing the probes into pools. E.g., a computer-aided method can be used to select or locate a group of nucleotide probes in which the frequency of repeated identification regions is minimized or reduced to zero. If there are no repeats there need be only one pool. If there is non-uniqueness, i.e., if an identification region is repeated, uniqueness can be 20 obtained by placing the nucleotide probes having the same identification region in separate pools.

25 In preferred embodiments the nucleotide probes are labeled, e.g., so that hybridization to the array of capture probes can be followed. The label can be a fluorescent label. The nucleotide probes of a first pool can be labeled with a first label and the nucleotide probes of a second pool can be labeled with a second label. If only one label is used then pools must be hybridized separately to the array of capture probes.

30 In preferred embodiments the capture probes are single stranded probes in an ordered array.

In preferred embodiments the capture probes have a structure comprising a double stranded region and a single stranded region in an ordered array.

26.

In preferred embodiments hybridization is detected by mass spectrophotometry, e.g., by MALDI-TOF mass spectrophotometry.

5 In preferred embodiments probes are selected for minimal crosshybridization with other probes.

In preferred embodiments the target or the selector probe has attached thereto a first member of a proximity detector pair and hybridization to the array allows the first member to be brought into proximity with a second member to provide a signal.

10

In preferred embodiments the target or a nucleotide probe has attached thereto a first member of a proximity detector pair and hybridization to the array allows the first member to be brought into proximity with a second member to provide a signal.

15

In preferred embodiments, the target is anchored to a substrate.

In methods described herein target or sample molecules can be amplified, e.g., by PCR, NASBA, RCA or other methods, prior to use.

20

Methods of U.S. 5,503,980 and or U.S. 5,631,134, both of which are hereby incorporated by reference can be used in methods of the invention, particularly, the array and array-related steps recited herein can use methods taught in these patents.

25

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Detailed Description

The drawings will first be briefly described.

30

#### Drawings

Fig. 1 is a diagram of steps 1 and 2 of an embodiment of the invention.

Fig. 2 is a diagram of steps 3 and 4 of an embodiment of the invention.

Figs. 3 is a diagram of steps 5 and 6 of an embodiment of the invention.

Figs. 4 is a diagram of a DNA sequencing method.

#### Detection of Polymorphisms

The figures describe an embodiment of the invention. In steps 1 and 2 the target 5 DNA is analyzed and nucleotide probes which will hybridize to polymorphism containing regions are selected. Methods for selecting nucleotide probes in terms of length and sequence, e.g., to optimize hybridization characteristics, e.g., to reduce or eliminate secondary structure are known in the art. The probes should be selected to minimize repetition of the identification regions. The identification regions are used to identify a 10 nucleotide probe and thus a polymorphism in the target. If necessary the probes can be divided into pools such that a pool has nucleotide probes all having unique identification regions. Thus, when a pool is hybridized to the array, the location of hybridization will identify a polymorphism.

Steps 3 and 4 show the coupling of the target to a solid support, hybridization a pool 15 of the probes to the target, and washing of the hybridization complex

Step 5 shows recovery of the hybridized probes.

Step 6 shows hybridization to the array of capture probes and detection of 20 hybridization.

Methods of U.S. 5,503,980 and or U.S. 5,631,134, both of which are hereby incorporated by reference can be used in methods of the invention, particularly, the array and array-related steps can use methods taught in these patents.

#### Differential Display Strategy-SNP (Single Nucleotide Polymorphism) and Mutation 25 Detection

This method can integrate the identifier probe methods, e.g., those which use pools of probes, with a differential display approach.

1. Obtain target DNA from control and test samples (i.e. normal and mutant or 30 disease).
2. Fragment target DNA, e.g., digest target DNA's with a restriction enzyme (e.g., a Type IIIs enzyme) to generate target DNA fragments with staggered ends and denature target DNA fragments.

28.

3. (optionally) De-phosphorylate the target DNA fragments.
4. Capture single strands of the fragmented DNA such that the single strands will be the Watson (or the Crick) of both test and control, i.e., both test and control are either, both Watson or both Crick. Test and control strands are isolated in different reaction vessels.
- 5 This step can also be used to capture only a subset of fragments from the digest, e.g., by recognizing unique ends on the restriction fragments. This produces two separate samples: the Watson (or Crick) strand from the test DNA and the same strand from the control DNA. Each of these is coupled to a substrate.
5. Add probe pools, as described herein, to the anchored test DNA strand.
- 10 6. Isolate the probes that hybridized to the test DNA strand and add these probes to the anchored control DNA strand under the same conditions as were used in the previous step. This sequesters, by hybridization to control DNA, probes which recognize both test and control sequences. The control DNA should be present in excess, e.g., 10X excess.
7. Collect the bound and unbound probes in separate fractions (A & B) and apply ~  
15 each fraction to a separate ordered array, e.g., positional sequencing by hybridization array (PSBH array) (1 & 2). A change in a signal at an array position indicates which probe is present.
8. The unbound probes (fraction B) that are hybridized to array 1 represent those SNPs or mutations that are present in the test sample but not in the control sample.
- 20 9. The bound probes (fraction A) that are ligated to the PSBH array 2 represent those SNPs or mutations that are common to both the control and the test samples.

#### Differential Display Strategy-RNA, EST and cDNA Detection

This method can integrate the identifier probe methods, e.g., those which use pools  
25 of probes, to analyze diverse samples.

1. Provide control and test samples (e.g., normal and mutant or disease). Normalize the concentration of the two samples.
2. In the case of an mRNA sample, capture the mRNA on a solid support using anchored polyA oligos or an ordered array, e.g., a PSBH array. If EST's or cDNA's are

29.

used, capture the ESTs and cDNAs using an ordered array, e.g., a PSBH array. Prepare at least two identical arrays for each sample.

3. Hybridize probe pools described herein to the anchored test mRNA (or cDNA or EST) and collect, in separate fractions, both the bound and unbound probes. Repeat this 5 step if necessary.

4. Hybridize the bound probes to the array of control using conditions identical to those in the previous step.

5. Collect the bound and unbound probes in separate fractions (A & B) and apply each fraction to separate arrays (1 & 2). The presence or absence of a signal at each array 10 position indicates which probes are present.

6. The unbound probes (fraction B) that are ligated array 1 represent those mRNA or cDNA species that are present in the test sample but not in the control sample.

7. The bound probes (fraction A) that are ligated to array 2 represent those mRNA or cDNA species that are common to both the control and the test samples.

15

#### Sequencing Method

1. Provide a probe having a first region complementary to a first region on a target nucleic acid, a second region complementary to a second region on the target nucleic acid, and optionally a third region disposed between. The first region on the target nucleic acid 20 and the second region on the target nucleic acid are separated by a region to be sequenced.

2. Hybridize the probe from step 1 to the target nucleic acid, thereby forming a single strand loop including the region of the target nucleic acid to be sequenced.

3. Optionally, label the target DNA or the probe.

4. Create breaks, e.g., random breaks, in the single strand loop, for example by 25 cleaving with an enzyme which cuts single strand nucleic acid.

5. Isolate cleavage fragments from step 4 and sequence them, e.g., by applying them to an ordered array, e.g., a PSBH array. Label present on either a fragment of the target nucleic acid (or on the probe fragment (if probe labelling is used the probe is cleaved in the third region) which is bound to the target nucleic acid fragment) allows for identification.

30.

DNA Structural Change Analysis

Insertions and deletions (or other events which alter chromosomal sequence) can be detected using proximity-label probes, e.g., FRET, HTRF, or LOCI probes. Target nucleic acid, e.g., fragments of genomic DNA generated by restriction digestion, are attached to a substrate in an array format (e.g., a Cantor array format). The target is denatured to form ssDNA (or ssDNA is generated using PCR off genomic DNA with one probe containing a linker molecule). A pool of probes, with first and second members of a proximity pair, e.g., donor and acceptor fluorophors, on separate probes, can be added to the genomic DNA.

The probes recognize various, non-overlapping, regions of the genomic DNA such that some are in close enough proximity to generate a signal, e.g., to allow FRET, HTRF, or LOCI to occur between the donor and acceptor fluorophors. Two probes that are adjacent or nearby in space will create a signal using FRET, HTRF, or LOCI that can be detected by the appropriate instrumentation. A signature pattern of, e.g., fluorescence or chemiluminescence, will be detected from the array using this approach. Insertions and deletions within the population of genomic DNA may disrupt the pattern of probe fluorescence or chemiluminescence on the array. In a variation of this approach, competition between probes (i.e. one of which contains a donor fluorophor or each containing a unique fluorophor) with overlapping binding sites can be used to detect different sequences.

SNP Ratio Analysis

This is a method for determining the presence of a single nucleotide polymorphism (SNP) marker that is based on comparing the ratio of nucleotides in a position in the genome between a population of control or "normal" genomic DNA and experimental or "disease" DNA. The concentration of genomic DNA extracted from each individual in a sample population is determined in order to normalize the concentration of the genomic DNA samples in the pool. A probe pool described herein is added to the immobilized genomic DNA, and the ratio of the four possible nucleotides for the SNP are determined for

31.

the population using multiplex labeled probes or multiple probe pools with a single label. This ratio is examined for a wide variety of SNPs. The SNP ratios observed for one population can be compared to another population. This can be combined with the differential display methods described herein.

5

#### Positional Arrays

Positional arrays suitable for the present invention include high and low density arrays on a two dimensional or three dimensional surface. Positional arrays include nucleic acid molecules, peptide nucleic acids or high affinity binding molecules of known sequence attached to predefined locations on a surface. Arrays of this nature are described in numerous patents which are incorporated herein by reference, Cantor US 5,503,980 (referred to herein as a Cantor array); Southern EP 0373 203 B1; Southern 5,700,637 and Deugau 5,508,169. The density of the array can range from a low density format, e.g., a microtiter plate, e.g., a 96- or 384- well microtiter plate, to a high density format, e.g. 1000 molecules /cm<sup>2</sup>, as described in Fodor US 5,445,934.

The surface on which the arrays are formed can be two dimensional, e.g. glass, plastic, polystyrene, or three dimensional, e.g. polymer gel pads, e.g. polyacrylamide gel pads of a selected depth, width and height. See, e.g., gel pads described in U.S. 5,552,270, hereby incorporated by reference.

20

The target or probes should bind to (and can be eluted from) the array at a single temperature. This can be effected by manipulating the length or concentration of the array or nucleic acid which hybridizes to it, by manipulating ionic strength or by providing modified bases.

25

Polymerase based extension of the capture probe and target or selector probe hybrid can be used to detect hybridization and may increase accuracy. The capture probes can be peptide nucleic acids in an array. Probes in an array can be: nucleic acids or peptide nucleic acids; hairpin-loop structures. Each capture probe can be in a discrete location on a support,

32.

or in a discrete location in a microtiter plate. A capture probe can be disposed in a gel pad on a support, e.g., a support comprising more than one gel pads.

### Proximity Methods

5 Proximity methods include those methods whereby a signal is generated when a first member and second member of a proximity detection pair are brought into close proximity. A "proximity detection pair" will have two members, the first member, e.g., an energy absorbing donor or a photosensitive molecule and the second member, e.g., an energy absorbing acceptor or a chemiluminescer particle. When the first and second members of  
10 the proximity detection pair are brought into close proximity, a signal is generated.

Examples of proximity methods include the following:

#### Fluorescence resonance energy transfer (FRET)

15 Fluorescence resonance energy transfer (FRET) is based on a donor fluorophore that absorbs a photon of energy and enters an excited state. The donor fluorophore transfers its energy to an acceptor fluorophore when the two fluorophores are in close proximity by a process of non-radiative energy transfer. The acceptor fluorophore enters an excited state and eliminates the energy via radiative or non-radiative processes. Transfer of energy from  
20 the donor fluorophore to acceptor fluorophore only occurs if the two fluorophores are in close proximity.

#### Homogeneous time resolved fluorescence (HTRF)

Homogeneous time resolved fluorescence (HTRF) uses FRET between two  
25 fluorophores and measures the fluorescent signals from a homogenous assay in which all components of the assay are present during measurement. The fluorescent signal from HTRF is measured after a time delay, thereby eliminating interfering signals. One example of the donor and acceptor fluorophores in HTRF include europium cryptate [(Eu)K] and XL665, respectively.

33.

Luminescent oxygen channelling assay (LOCI)

In the luminescent oxygen channelling assay (LOCI), the proximity detection pairs includes a first member which is a sensitizer particle that contains phthalocyanine. The 5 phthalocyanine absorbs energy at 680nm and produces singlet oxygen. The second member is a chemiluminescer particle that contains olefin which reacts with the singlet oxygen to produce chemiluminescence which decays in one second and is measured at 570nm. The reaction with the singlet oxygen and the subsequent emission depends on the proximity of the first and second members of the proximity detection pair.

10

Other embodiments are within the following claims.

What is claimed is:

34.

**Claims**

1. A method of analyzing a polynucleotide sequence in a sample comprising:  
providing a sample which includes a polynucleotide sequence to be analyzed and at  
5 least one extraneous nucleic acid molecule;  
providing a plurality of selector probes wherein each selector probe is  
complementary to a region of the polynucleotide sequence to be analyzed and is not  
complementary to the extraneous nucleic acid;  
contacting the sample with the plurality of selector probes under conditions which  
10 allow hybridization of the polynucleotide sequence to be analyzed to a selector probe but  
which do not allow hybridization of the extraneous nucleic acid to hybridize to a selector  
probe;  
allowing a nucleic acid molecule to be analyzed to hybridize with a selector probe;  
selecting a product of that hybridization, e.g., the selector probe, the nucleic acid to  
15 be analyzed, or both;  
providing an array of a plurality of capture probes, wherein each of the capture  
probes is positionally distinguishable from other capture probes of the plurality on the array,  
and wherein each positionally distinguishable capture probe includes a unique (i.e., not  
repeated in another capture probe) region complementary to the plurality of selector probes;  
20 hybridizing the selected nucleic acid molecule (which does not include the  
extraneous nucleic acid molecule) with the array of capture probes, thereby detecting or  
identifying a selected nucleic acid molecule which bound to the polynucleotide sequence  
and thereby analyzing the polynucleotide sequence.
- 25 2. The method of claim 1 wherein the polynucleotide sequence is: a DNA molecule:  
all or part of a known gene; wild type DNA; mutant DNA; a genomic fragment, particularly  
a human genomic fragment; a cDNA, particularly a human cDNA.
- 30 3. A method of detecting the presence of, or identifying, a genetic event comprising:  
providing a target nucleic acid having a genetic event;  
providing a plurality of nucleotide probes wherein each probe:  
(a) is complementary to a region immediately 5' to the site of the genetic  
event and complementary to a region immediately 3' to the site of the genetic  
event ,

35.

(b) has an identification region, e.g., a 5' terminal region or a 3' terminal region, which identification region is not repeated in another probe in the plurality, i.e., it is unique in the plurality, and is complementary to the target sequence, and

5 (c) wherein the nucleotide at the position corresponding to the site of the first genetic event can be chosen from A, T, G, or C,

wherein the plurality of nucleotide probes includes a first probe having a first base at the position corresponding to the site of the genetic event and a second probe having a second (i.e., different from the first base) base at the position corresponding to the site of the 10 genetic event;

contacting the target with the plurality of probes under conditions wherein only probes having complementarity with the target nucleic acid will hybridize; thereby analysing the target nucleic acid.

15 4. The method of claim 3 wherein the identification region can allow identification of the identity of a base at the genetic event.

5. The method of claim 3, further comprising:

providing an array of a plurality of capture probes, wherein each of the capture 20 probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region complementary to the identification region of a probe of the plurality of nucleotide probes;

hybridizing the nucleotide probes which hybridize to the target to the array of 25 capture probes, thereby detecting or identifying a nucleotide probe which bound to the target and thereby identifying the genetic event.

6. The method of claim 3 wherein the identification region is a naturally occurring sequence in the target molecule.

30

7. A method of detecting the presence of, or identifying, one or more genetic events, comprising:

36.

(1) providing a target nucleic acid having a first genetic event, e.g., a polymorphism, and a second genetic event;

5 (2) providing a first plurality of nucleotide probes wherein each probe:

(a) is complementary to a region immediately 5' to the site of first the first polymorphism and complementary to a region immediately 3' to the site of the genetic event,

(b) has an identification region, e.g., a 5' terminal region or a 3' terminal region, which is complementary to the target sequence, and

10 (c) wherein the nucleotide at the position corresponding to the site of the first polymorphism can be chosen from A, T, G, or C,

(3) providing a second plurality of nucleotide probes wherein each probe:

15 (a) is complementary to a region immediately 5' to the site of the second polymorphism and complementary to a region immediately 3' to the site of the second genetic event,

(b) has an identification region, e.g., a 5' terminal region or a 3' terminal region, which is complementary to the target sequence, and

20 (c) wherein the nucleotide at the position corresponding to the site of the second polymorphism can be chosen from A, T, G, or C,

(4) optionally, organizing the first and second plurality of probes into one or more pools, such that within a pool, an identification region is not repeated in the pool, i.e., it is unique in the pool, and

25

(5) contacting the target with the first and second plurality of nucleotide probes under conditions wherein only probes having complementarity with the target will hybridize;

30 thereby analysing the target nucleic acid. The identification region can allow identification of the identity of a base at the genetic event.

8. The method of claim 7, further comprising a particularly preferred embodiment the method further includes:

- (6) providing an array of a plurality of capture probes, wherein each of the plurality of capture probes is positionally distinguishable from other capture probes of the plurality on the array, and wherein each positionally distinguishable capture probe includes a unique (i.e., not repeated in another capture probe) region complementary to the identification region of a probe of the first and second plurality of nucleotide probes;
- (7) hybridizing nucleotide probes which hybridize to the target to the array, thereby detecting or identifying a nucleotide probe which bound to the target and thereby identifying the polymorphism, provided that in a hybridization all of the nucleotide probes have a unique identification region, e.g., they are all from a single pool.
- 10 9. A method of analyzing a target nucleotide sequence in a sample comprising:  
providing a target nucleic acid;  
hybridizing to the target nucleic acid a first probe having attached thereto a first  
15 member of a proximity detection pair;  
hybridizing to the target nucleic acid a second probe having attached thereto a second member of a proximity detection pair; and  
determining if a signal is generated by the proximity detection pair.
- 20 10. A method of analyzing a polynucleotide sequence in a sample comprising:  
providing a target nucleic acid;  
hybridizing to the target nucleic acid a first probe having attached thereto a first member of a proximity detection pair;  
hybridizing to the target nucleic acid a second probe having attached thereto a  
25 second member of a proximity detection pair, wherein said second probe hybridizes to a region which contains a genetic event, e.g., a polymorphism, e.g., a single nucleotide polymorphism, and wherein hybridization of the second probe is determined by the presence or absence of the genetic event;  
determining if a signal is generated by the proximity detection pair.
- 30 11. A method of analyzing a polynucleotide sequence in a sample comprising:  
providing a sample which includes a polynucleotide sequence to be analyzed and at least one extraneous nucleic acid molecule;

providing a plurality of selector probes wherein each selector probe is complementary to a region of the polynucleotide sequence to be analyzed and is not complementary to the extraneous nucleic acid,

5 contacting the sample with the plurality of selector probes under conditions which allow hybridization of the polynucleotide sequence to be analyzed to a selector probe but which do not allow hybridization of the extraneous nucleic acid to hybridize to a selector probe;

allowing a nucleic acid molecule to be analyzed to hybridize with a selector probe;

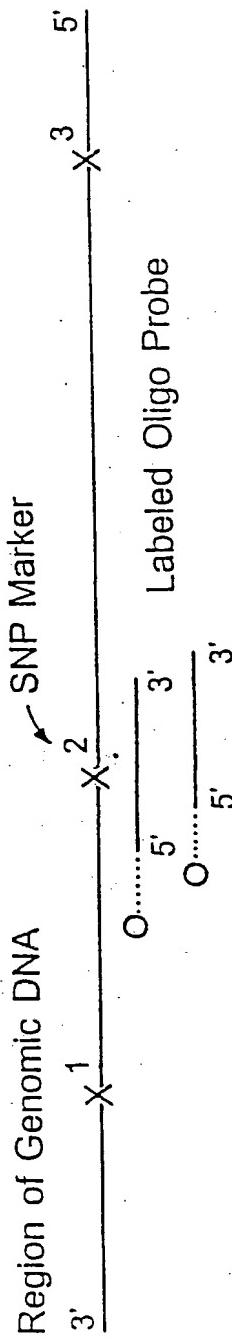
selecting the selector probes that hybridise to the polynucleotide sequence;

10 detecting or identifying the selector probes which bound to the polynucleotide sequence and thereby analyzing the polynucleotide sequence.

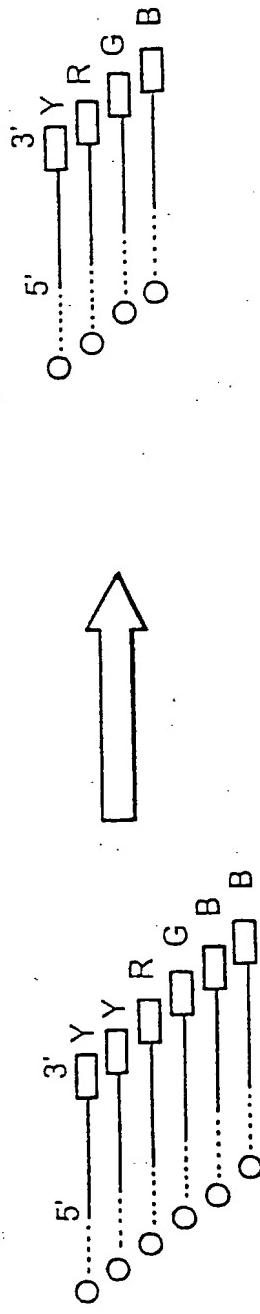
1/4

### The SNaP Platform

1. Select oligo probes that are optimized for hybridizing to SNP sites in genomic DNA.



2. Organize the oligo SNP probes for interrogating genomic DNA into multiple pools so that the 3' five bases of each probe in the pool are different from those of the other probes in the same pool.



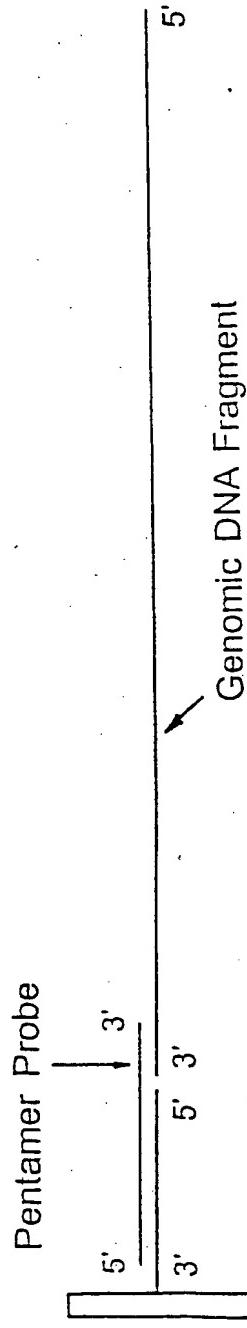
Selection to eliminate non-unique end regions  
i.e., to remove identical 3' or 5' sequence.

**FIG. 1**

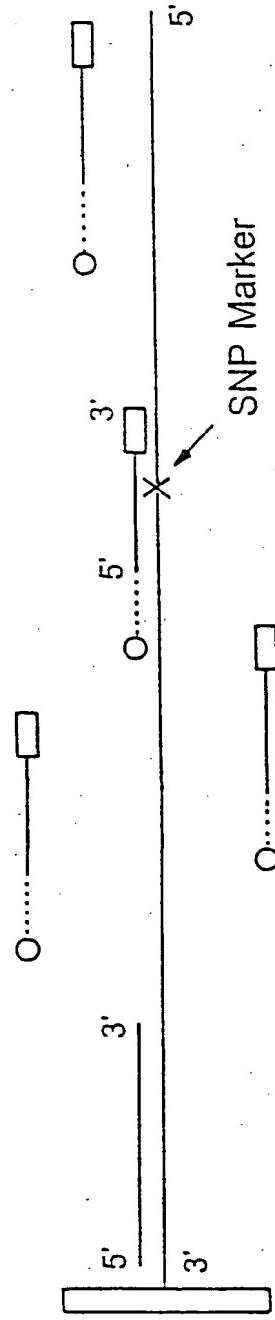
2/4

### The SNaP Platform

3. Ligate denatured and restriction digested genomic DNA to a solid support.



4. Hybridize one of the pools of oligo SNP probes to the anchored genomic DNA and wash away unhybridized probes.

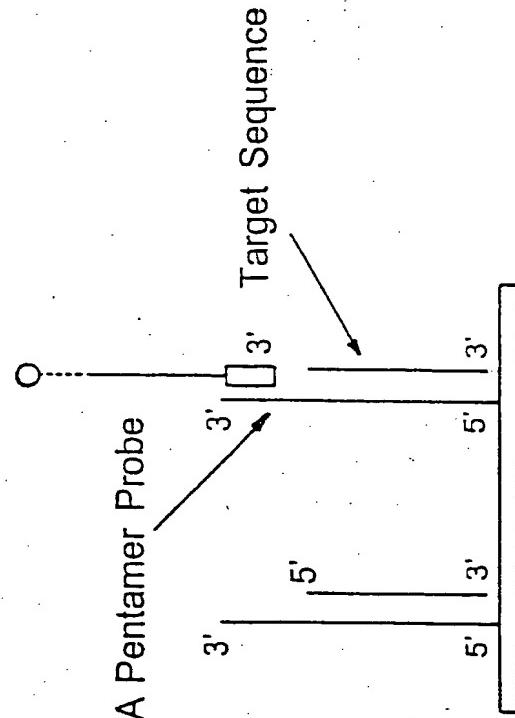
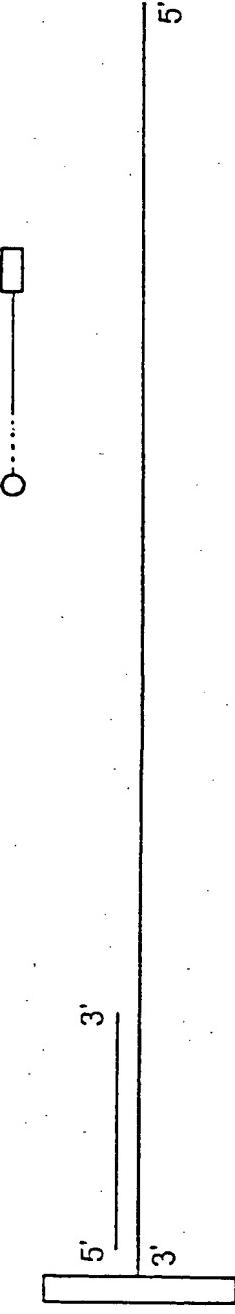


**FIG. 2**

3/4

## The SNaP Platform

### 5. Denature the oligo SNP probe from the genomic DNA.

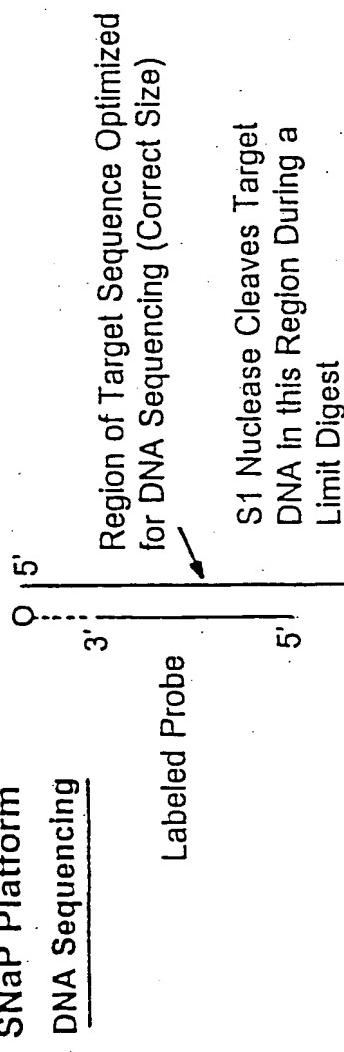


6. Capture the oligo SNP probe on an array of dsDNA probes containing a random, determinable five base 3' extension with a ligation reaction (PSBH Approach). Repeat steps 4, 5 and 6 with the other oligo SNP probe pools.

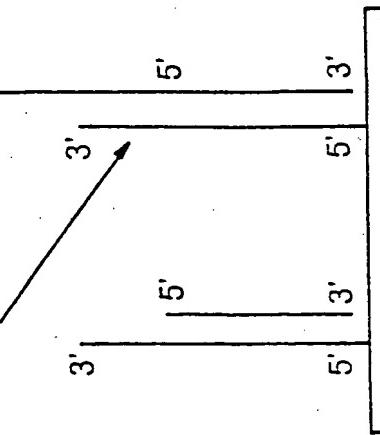
**FIG. 3**

4/4

**SNaP Platform**  
**DNA Sequencing**



ssDNA Pentamer Probe Captures Target DNA by Ligation



The DNA sequences of the 3' Ends of the Fragments Generated by the S1 Nuclease Digest are Determined Using the Cantor Array

**FIG. 4**

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/25664

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/66; C12P 19/34; C07H 21/00  
 US CL : 435/6, 91.1; 536/24.31,24.32

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 91.1; 536/24.31,24.32

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	GUO, Z. et al. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. Nucleic Acids Research. 1994, Vol. 22, No. 24, pages 5456-5465, see entire document especially abstract.	1,2
Y	US 5,503,980 A (CANTOR et al) 02 April 1996 (04/02/96), see entire document, especially column 8, lines 18-36.	1-11
A,P	US 5,753,439 A (SMITH et al) 19 November 1998 (11/19/98), see entire document, especially abstract.	1-11



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
* "A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
* "E" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
* "L" document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
* "O" document referring to an oral disclosure, use, exhibition or other means		
* "P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

15 JANUARY 1999

Date of mailing of the international search report

08 FEB 1999

Name and mailing address of the ISA/US  
 Commissioner of Patents and Trademarks  
 Box PCT  
 Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

GARY JONES

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/25664

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SAIKI, R.K. et al. Genetic analysis of amplified DNA with immobilized sequence -specific oligonucleotide probes. PNAS. August 1989, Vol. 86, No. 16, see pp. 6230-6234, see entire document.	1-11
Y	JUNG, P.M. et al. Energy transfer of nucleic acid products. Prog. Bio. Optics. 08-10 February 1995, Vol. 2386, see pages 118-127.	9-11
Y	MATTHEWS, J.A. et al. Analytical Strategies for the use of DNA probes. Anal. Bio. February 1988, Vol.169, pages 1-25,especially page 18.	9-11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/25664

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, STN, WPIDS, BIOSIS, MEDLINE, CANCERLIT, BIOTECHS, LIFESCI, CAPLUS, EMBASE  
search terms:, probe , oligonucleotide, array, ligation, complementary, sequence

